

Allier recherche – diffusion de sa recherche – enseignement : est-ce possible?

François Husson

<https://husson.github.io>

Unité de mathématiques appliquées, Agrocampus Ouest, Rennes

PROGEDO-Loire - Data SHS – 13 décembre 2019

Présentation

- Recherche : données manquantes, analyse de données, tableaux multiples
- Enseignement : cursus d'ingénieur, master *science des données*
- MOOC en analyse de données et MOOC en Sensométrie
- Formation continue : statistique avec R, analyse de données



2018



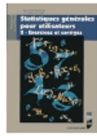
2nd ed: 2017

1st ed: 2011



2nd ed: 2016

1st ed: 2009



2nd ed: 2013

1st ed: 2005



2013



3rd ed: 2012

2nd ed: 2010

1st ed: 2008



2012

Packages:

FACTOMINER^R - missMDA - SensoMineR^R - Factoshiny -

FactoInvestigate - RcmdrPlugin.FactoMineR

Plan

Introduction



Diffusion

Enseignement

Introduction

Construire un package permet de :

- proposer de nouvelles méthodes ou méthodologies statistiques
- partager son travail avec la communauté scientifique
- faciliter la confrontation de méthodes
- rendre disponibles des jeux de données

Créer un package est chronophage : cela DOIT être bénéfique au concepteur du package ET à la communauté scientifique

Introduction

Construire un package permet de :

- proposer de nouvelles méthodes ou méthodologies statistiques
- partager son travail avec la communauté scientifique
- faciliter la confrontation de méthodes
- rendre disponibles des jeux de données

Créer un package est chronophage : cela DOIT être bénéfique au concepteur du package ET à la communauté scientifique

Questions à se poser avant la soumission du package

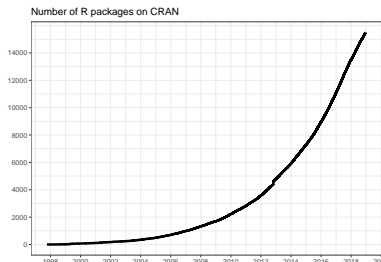
- Un package pour qui ? Pour quoi ? qq chercheurs du domaine (github, zip) ? audience large (bioconductor, CRAN) ?
- 1ère version du package peut être limitée mais bien faite
 - qq choix seront difficiles à modifier : nom package, fonctions, arguments
 - comprendre le fonctionnement : documenter fonctions, exemples, vignette

De plus en plus de packages

Nb de packages en déc 2019

- CRAN : 16 364
- Bioconductor : 1 903
- R-Forge : 2 116
- GitHub : 53 106

⇒ la visibilité d'un package est de + en + faible

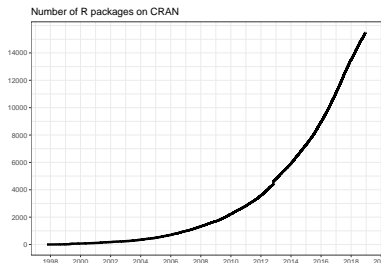


De plus en plus de packages

Nb de packages en déc 2019

- CRAN : 16 364
- Bioconductor : 1 903
- R-Forge : 2 116
- GitHub : 53 106

⇒ la visibilité d'un package est de + en + faible



Après la soumission du package

- "Faire vivre" le package : corriger coquilles, répondre aux questions, nouveaux développements, améliorer la programmation (Rcpp, parallélisation...)
- Construire des packages complémentaires
- Faire connaître le package

Plan

Introduction

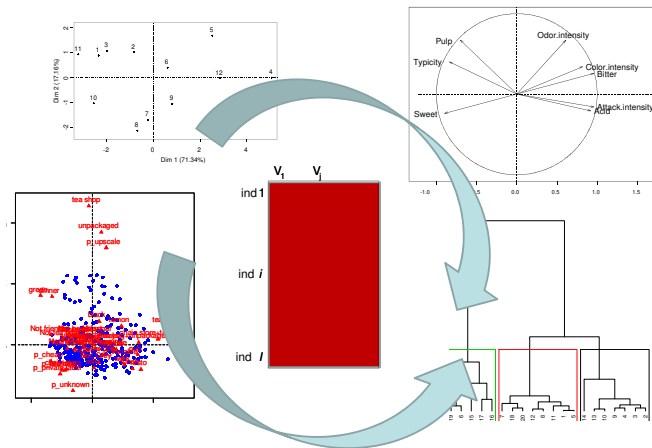


Diffusion

Enseignement



et les méthodes d'analyse de données



Objectifs :

- Descriptif - exploratoire : visualisation de données
- Synthèse - résumé de grands tableaux individus \times variables

FACTOMINER[®] en quelques mots

Le package

- propose des méthodes d'analyses factorielles et de classification (ACP, ana. des correspondances simples et multiple, ana. fact. de données mixtes, ana. fact. multiple, ana. fact. multiple hiérarchique, CAH)
- interface graphique en français et anglais (*Factoshiny*)
- gestion des données manquantes (*missMDA*)
- rapport automatisé (*FactoInvestigate*)
- aides à l'utilisateur (site internet, vidéos, livres, MOOC)

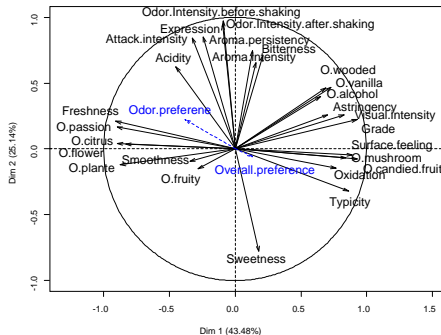
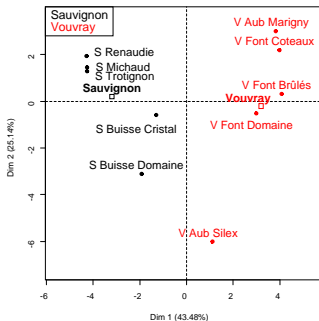
Les données vins : description par des experts

- 10 individus : vins blancs du Val de Loire
- 27 variables quantitatives : descripteurs sensoriels
 - mais aussi 2 variables d'appréciation
 - et 1 variable qualitative : label des vins (Vouvray - Sauvignon)

	O.fruity	O.passion	O.citrus	...	Sweetness	Acidity	Bitterness	Astringency	Aroma.intensity	Aroma.persistency	Visual.intensity	Odor.preference	Overall.preference	Label
S Michaud	4.3	2.4	5.7	...	3.5	5.9	4.1	1.4	7.1	6.7	5.0	6.0	5.0	Sauvignon
S Renaudie	4.4	3.1	5.3	...	3.3	6.8	3.8	2.3	7.2	6.6	3.4	5.4	5.5	Sauvignon
S Trotignon	5.1	4.0	5.3	...	3.0	6.1	4.1	2.4	6.1	6.1	3.0	5.0	5.5	Sauvignon
S Buisse Domaine	4.3	2.4	3.6	...	3.9	5.6	2.5	3.0	4.9	5.1	4.1	5.3	4.6	Sauvignon
S Buisse Cristal	5.6	3.1	3.5	...	3.4	6.6	5.0	3.1	6.1	5.1	3.6	6.1	5.0	Sauvignon
V Aub Silex	3.9	0.7	3.3	...	7.9	4.4	3.0	2.4	5.9	5.6	4.0	5.0	5.5	Vouvray
V Aub Marigny	2.1	0.7	1.0	...	3.5	6.4	5.0	4.0	6.3	6.7	6.0	5.1	4.1	Vouvray
V Font Domaine	5.1	0.5	2.5	...	3.0	5.7	4.0	2.5	6.7	6.3	6.4	4.4	5.1	Vouvray
V Font Brûlés	5.1	0.8	3.8	...	3.9	5.4	4.0	3.1	7.0	6.1	7.4	4.4	6.4	Vouvray
V Font Coteaux	4.1	0.9	2.7	...	3.8	5.1	4.3	4.3	7.3	6.6	6.3	6.0	5.7	Vouvray

- Quels vins se ressemblent ? Peut-on faire des groupes de vins ?
- Comment caractériser un vin ?
- Quels descripteurs se ressemblent ?

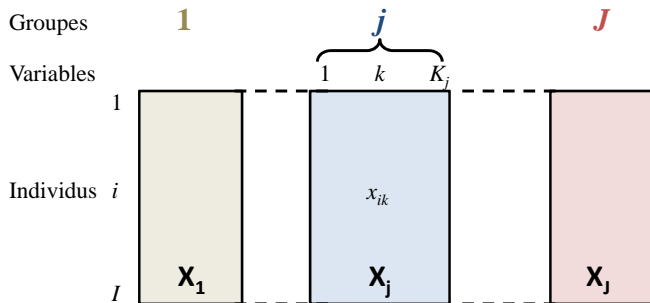
Représentation des individus et des variables



⇒ Utilisation d'information supplémentaire

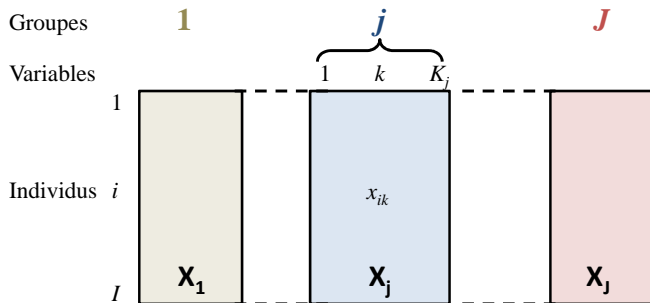
- la variable qualitative *cépage*
- les variables quantitatives d'*appréciation*

L'Analyse Factorielle Multiple (AFM)



Exemples avec des variables **quantitatives et/ou qualitatives**
et/ou des tableaux de contingence :

L'Analyse Factorielle Multiple (AFM)



Exemples avec des variables **quantitatives et/ou qualitatives**
et/ou des tableaux de contingence :

- enquête *mieux vivre* par pays (22 indicateurs de 5 domaines)
- tableau pays \times indicateurs économique, sur plusieurs années
- questionnaire avec échelles de likert et questions qualitatives
- analyse textuelle d'un mouvement social par les journaux, à plusieurs dates

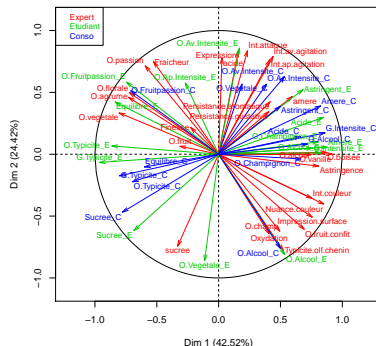
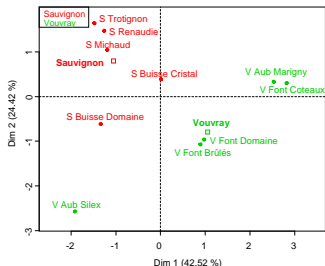
Description sensorielle de vins : comparaison de jurys

- description sensorielle de 3 jurys : œnologue, conso., étudiant

	Expert (27)	Conso (15)	Etudiant (15)	Appréciation (60)	Cépage (1)
Vin 1					
Vin 2					
...					
Vin 10					

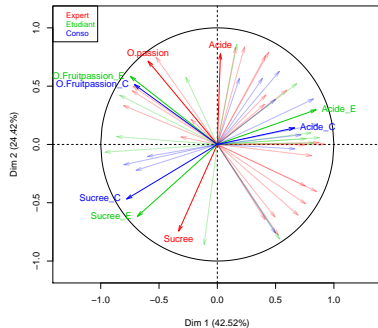
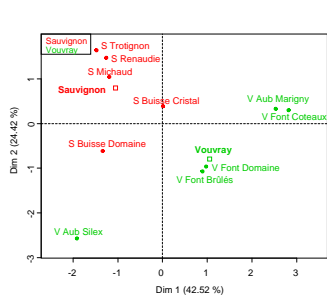
- Comment caractériser les vins ?
- Les vins sont-ils décrits de la même façon par les différents jurys ? Y-a t'il des spécificités par jury ?
- Peut-on comparer les typologies des vins d'un jury à l'autre ?

Représentation des individus et des variables



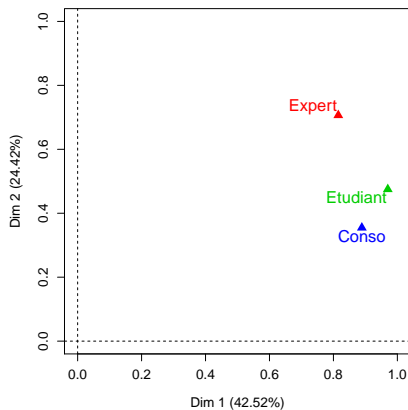
Même interprétation qu'en ACP

Représentation des individus et des variables



Même interprétation qu'en ACP

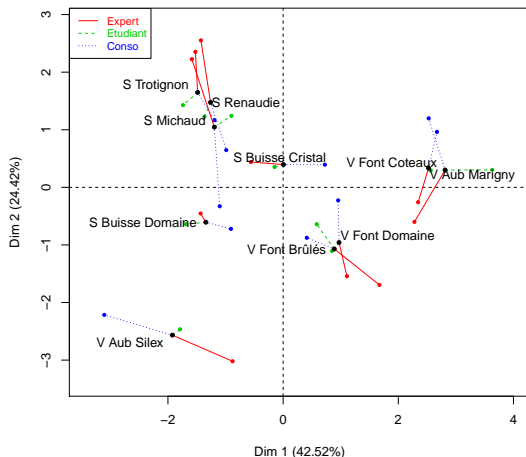
Représentation des groupes



- 1ère dimension commune à tous les groupes
- 2ème dimension due au groupe Expert
- 2 groupes sont proches quand ils induisent la même structure

⇒ Ce graphe fournit une comparaison synthétique des groupes
⇒ Les positions relatives des individus sont-elles similaires d'un groupe à l'autre ?

Représentation des points partiels



- Point partiel = représentation d'un individu vu par un groupe
- Un individu est au barycentre de ses points partiels
- Un individu est homogène si ses points partiels sont proches

Un exemple en linguistique

- | | |
|---------------------------|---|
| - Aragon (23 textes) : | FeuJoie, Perpétuel, Destinées, Snark, Peinture, ... |
| - Balzac (49 textes) : | <i>Chouans, Physiologie, Vendetta, Gobseck, ...</i> |
| - Corneille (34 textes) : | <i>Mélite, Clitandre, Veuve, Galerie, Suivante, ...</i> |
| - ... | |



Un exemple en linguistique

- Aragon (23 textes) : FeuJoie, Perpétuel, Destinées, Snark, Peinture, ...
- Balzac (49 textes) : Chouans, Physiologie, Vendetta, Gobseck, ...
- Corneille (34 textes) : Méliste, Clitandre, Veuve, Galerie, Suivante, ...
- ...

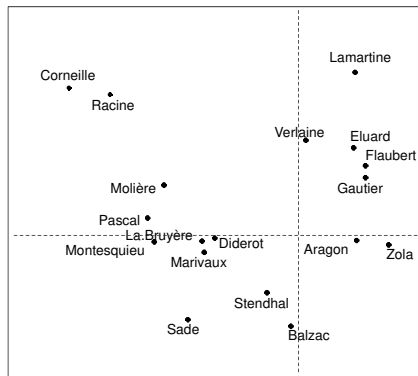
On conserve les
mots cités au
moins 100 fois

978 mots



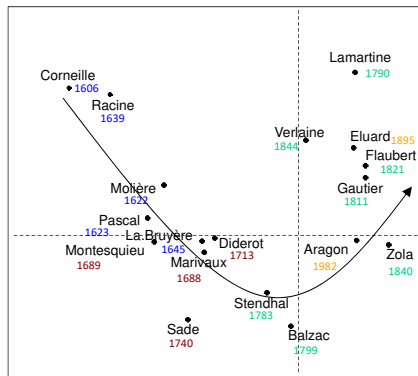
accord	264	0	88	44	...
affaire	1029	2040	74	154	...
âge	545	629	92	108	
ah	219	0	0	0	
air	2093	2009	95	191	
allemagne	366	0	0	0	
allemand	476	0	0	0	
amant	303	760	566	0	
âme	478	2190	1101	240	
ami	1090	2583	307	407	
amour	1374	3286	1791	167	
an	1812	3009	112	182	
anglais	315	0	0	0	
. . .					

Un exemple en linguistique : l'analyse des correspondances



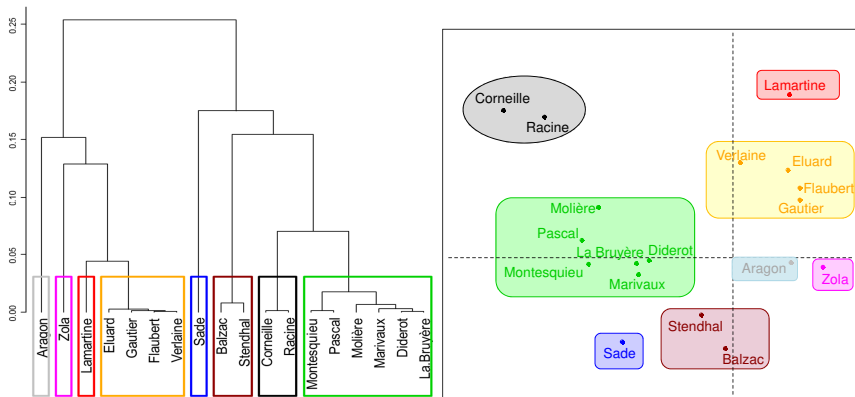
Avec l'AFC, les auteurs sont d'autant plus proches qu'ils emploient les mots dans les mêmes proportions, i.e. qu'ils s'intéressent aux mêmes sujets et ont les mêmes préoccupations

Un exemple en linguistique : l'analyse des correspondances



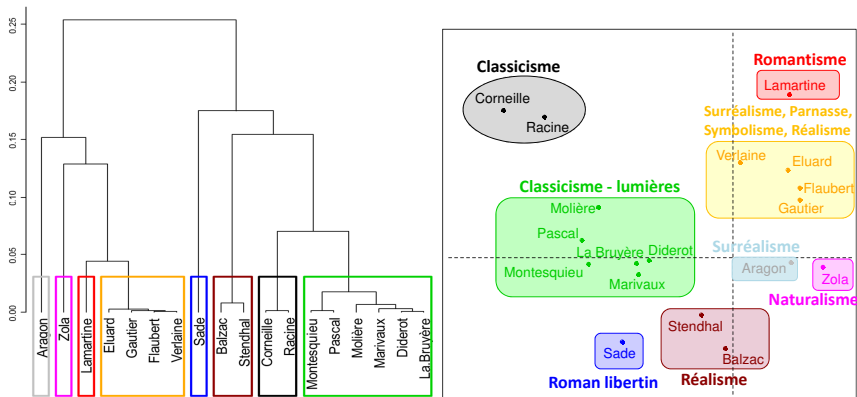
- Corneille et Racine sont proches et très éloignés de Zola. Ce sont 2 auteurs classiques du 17ème tandis que Zola est un naturaliste du 19ème
- Évolution du vocabulaire selon les siècles

Un exemple en linguistique : caractérisation des classes



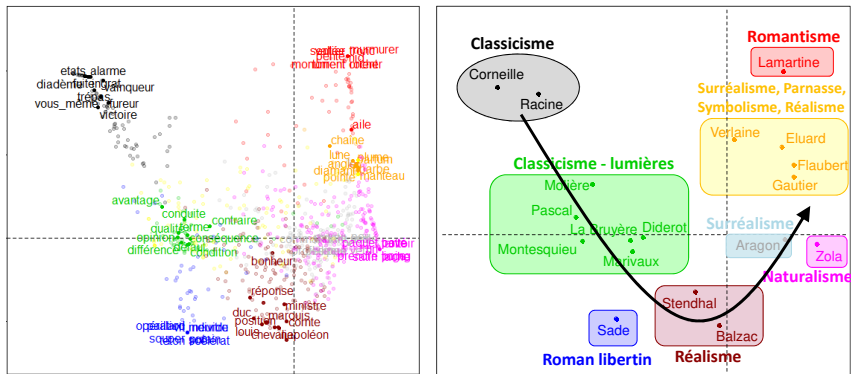
- La classification retrouve des classes d'auteurs connues

Un exemple en linguistique : caractérisation des classes



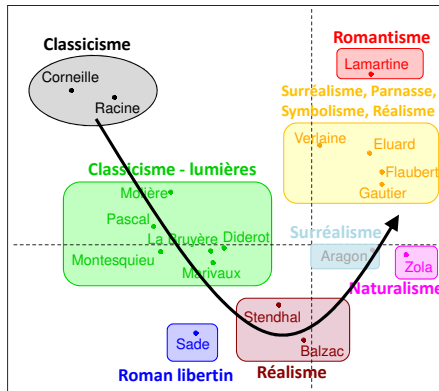
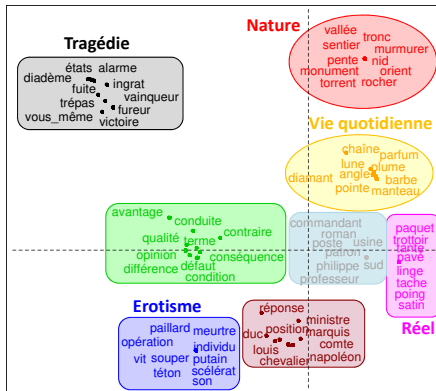
- Stendhal et Balzac (réalistes) sont très éloignés de Lamartine (romantique). On retrouve ici que les auteurs réalistes ont un point commun : s'éloigner des excès romantiques !
- Points communs naturalistes / réalistes : montrer la société telle qu'elle est, le roman devient le miroir de la société

Un exemple en linguistique : interprétation ...



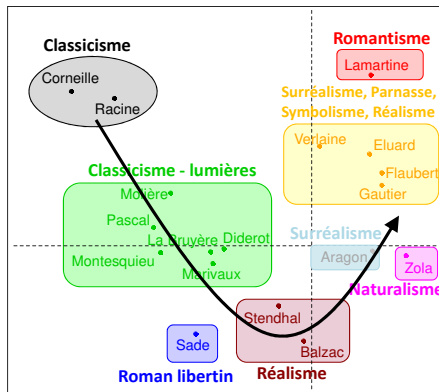
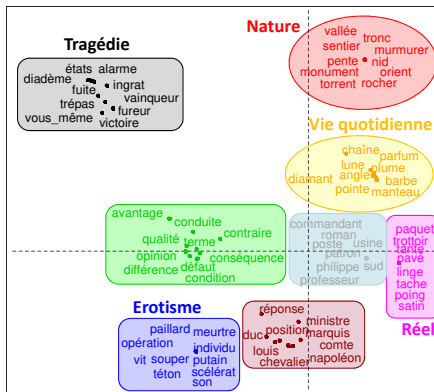
Les mots permettent de caractériser les sujets de prédilection des auteurs et les courants littéraires

Un exemple en linguistique : interprétation ...



- Le naturalisme est la suite logique du réalisme : le naturalisme montre le milieu où vit le protagoniste pour expliquer son comportement de façon "scientifique"
- Évolution du vocabulaire selon les courants littéraires

Un exemple en linguistique : interprétation ...



- Le naturalisme est la suite logique du réalisme : le naturalisme montre le milieu où vit le protagoniste pour expliquer son comportement de façon "scientifique"
- Évolution du vocabulaire selon les courants littéraires

Une vidéo pour en savoir plus sur cet exemple

missMDA : un package pour gérer les données manquantes

Variables

1 j p

Individus

1

i

n

The diagram shows a light blue rectangular matrix representing a dataset. The horizontal axis is labeled 'Variables' with indices 1, j, and p. The vertical axis is labeled 'Individus' with indices 1, i, and n. The matrix contains several question marks '?' scattered throughout, representing missing data points. For example, in the first row (i=1), there are missing values at columns 1, 2, 4, 5, and 6. In the i-th row, there are missing values at columns 2, 3, 4, 5, and 6. In the n-th row, there are missing values at columns 1, 3, 4, 5, and 6.

Etude et implémentation des données manquantes en ACP, AFC, ACM, AFM et AFDM

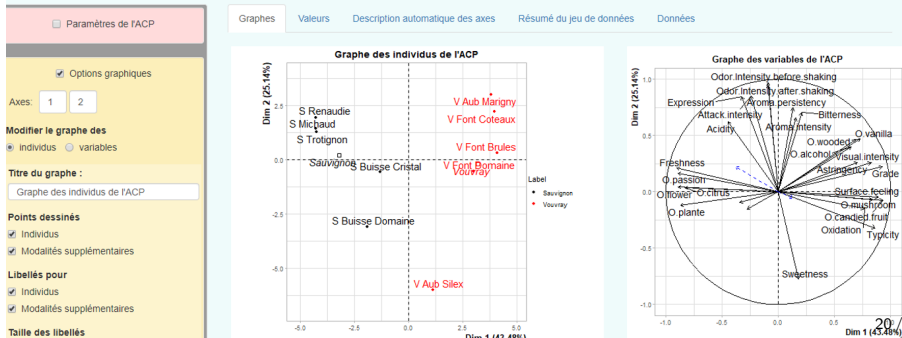
1. Imputation par les méthodes d'analyse de données
2. Analyse du tableau imputé

Factoshiny : un package pour des graphiques interactifs

- Réaliser des analyses sans besoin de maîtriser le code
- Visualisation en temps réel des modifications apportées

```
library(Factoshiny)
vins <- read.table("https://husson.github.io/img/vins_expert.csv",
                  header=TRUE, sep=";", row.names=1)
res <- Factoshiny(vins)      ## analyse factorielle sur les données
res <- Factoshiny(res.pca)   ## graphe sur un objet résultat de FactoMineR
res2 <- Factoshiny(res)     ## objet résultat de Factoshiny
```

ACP sur le jeu de données vins



FactoInvestigate : un package pour des rapports automatisés

Propose une interprétation des résultats basée sur l'objet résultat

```
> res.pca <- PCA(MesDonnees, ...)
> library(FactoInvestigate)
> Investigate(res.pca)
```

Analyse en Composantes Principales

Jeu de données decathlon

Ce jeu de données contient 41 individus et 13 variables. 2 variables quantitatives sont fléchies, 1 variable qualitative est fléchée.

1. Observation d'individus extrêmes

L'analyse des graphes ne révèle aucun individu singulier.

2. Distribution de l'inertie

L'inertie des axes factoriels indique d'une part si les variables sont structurées et suggère d'autre part le nombre judicieux de composantes principales à étudier.

Les 2 premiers axes de l'ACP expliquent **80.09%** de l'inertie totale du jeu de données ; cela signifie que 80.09% de la variabilité totale du nuage des individus (ou des variables) est représentée dans ce plan. C'est un pourcentage assez important, et le premier plan représente donc, concrètement la variabilité contenue dans une grande part du jeu de données actif. Cette valeur est supérieure à la valeur référence de **37.73%**, la variabilité expliquée par ce plan est donc significative (cette inertie de référence est le quartile 0.95 de la distribution des pourcentages d'inertie obtenus en simulat 1000 jeux de données aléatoires de dimensions comparables sur la base d'une distribution normale).

Du fait de ces observations, il serait tout de même probablement préférable de considérer également dans l'analyse les dimensions supérieures ou égales à la troisième.

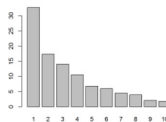
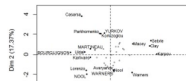


Figure 2 - Decomposition of the total inertia on the components of the ACP

Une estimation du nombre pertinent d'axes à interpréter suggère de restreindre l'analyse à la description des 3 premiers axes. Ces composantes révèlent un taux d'inertie supérieur à celle du quartile 0.95 de distributions aléatoires (34.74% contre 37.44%). Cette observation suggère que seuls ces axes sont porteurs d'une véritable information. En conséquence, la description de l'analyse sera restreinte à ces seuls axes.

3. Description du plan 1:2



<http://factominer.free.fr/reporting>

Plan

Introduction



Diffusion

Enseignement

Aides à l'utilisateur : site internet

- <http://factominer.free.fr> en anglais et en français
- exemples, aides sur les fonctions, références, etc.

FactoMineR : analyse de

factominer.free.fr/index_fr.html

Accueil Méthodes FactoMineR Enseignement MOOC, livres Améliorations graphiques Valeurs manquantes missMDA Rapport automatique Google group Plus

FACTOMINER^R

FactoMineR en quelques mots

FactoMineR est un package R dédié à l'analyse exploratoire multidimensionnelle de données (à la Française). Il a été développé et il est maintenu par François Husson, Julie Josse, Sébastien Lê, d'Agrocampus Rennes, et J. Mazet.

Pourquoi utiliser FactoMineR?

1. Il permet de mettre en oeuvre des méthodes analyses de données telles que l'analyse en composantes principales (ACP), l'analyse des correspondances (AC), l'analyse des correspondances multiples (ACM) ainsi que des analyses plus avancées.
2. Il permet l'ajout d'information supplémentaire telle que des individus et/ou des variables supplémentaires.
3. Il fournit un point de vue géométrique et de nombreuses sorties graphiques.
4. Il fournit de nombreuses aides à l'interprétation (description automatique des axes, nombreux indicateurs, ...).
5. Il peut prendre en compte diverses structures sur les données (structure sur les variables, hiérarchie sur les variables, structure sur les individus).
6. Beaucoup de matériels pédagogique (MOOC, livres, etc.) est disponible pour expliquer aussi bien les méthodes que la façon de les mettre en oeuvre avec FactoMineR.
7. Il gère les données manquantes avec missMDA (voir ici).

Menu de l'accueil

- Présentation de FactoMineR
- Nouvelles
- Installation de FactoMineR
- Comment citer FactoMineR?
- Historique de FactoMineR

Liens utiles

- Département de Mathématiques d'Agrocampus Rennes
- R Project
- CRAN

Aides à l'utilisateur : un Google group

- <https://groups.google.com/group/factominer-users/>
- possibilité de poser des questions et/ou répondre
- en français ou en anglais

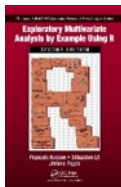
The screenshot shows a web browser window displaying the Google Groups page for 'FactoMineR users'. The URL in the address bar is <https://groups.google.com/forum/?hl=fr#forum/factominer-users>. The page header includes the Google logo, a search bar, and navigation links. Below the header, the group name 'FactoMineR users' is displayed, along with the number of subjects (32 sur 405) and a 'G+' icon. A table of recent posts is shown, with columns for subject, number of messages, and date.

Subject	Messages	Date
La 2ème session du MOOC "Analyse de données multidimensionnelles" débute le 1er mars	1 message	19 janv.
Nouveau module graphique	11	17/01/2015
Select the best sample using a reference	1	16 mai
salut	1	15 mai
PCAshiny scale.unit=F impossible ?	2	12 mai
ACM et questions à choix multiples (plusieurs modalités dans la même question)	6	2 mai
Plot CA neatly (1)	1	30 avr.
Interpreting MCA results	11	22 avr.
Installing FactoMineR on Linux (1)	1	21 avr.
Estimation of PC for MFA	4	21 avr.
General questions for FAMD	1	1 avr.

Aides à l'utilisateur : diffusion scientifique

- Suggéré le package dans un des [CRAN Task views](#)
- Exposés aux rencontres R et à useR !
- 2 articles dans J. of stat. software ([FactoMineR](#), [missMDA](#))
- 2 articles dans R journal ([CA-galt](#), [MFACT](#))
- Des livres :

Analyse de données avec R (2^e ed)



R pour la statistique et la science des données



Plan

Introduction



Diffusion

Enseignement

Aides à l'utilisateur : des vidéos

- disponibles depuis [ma page d'enseignement](#)
- disponibles depuis les aides de FactoMineR
- <https://www.youtube.com/HussonFrancois>
- \approx 60 vidéos en français ou anglais

YouTube

Rechercher

François Husson

6 147 abonnés

PERSONNALISER LA CHAÎNE

VERSION BÊTA DE YOUTUBE STUDIO

ACCUEIL VIDÉOS PLAYLISTS CHAÎNES DISCUSSION À PROPOS

Playlists créées

Multiple Factorial ANOVA 36

Classification ascendante 6

Clustering 5

Analyse Factorielle Multiple 8

Vidéos "J'aime"

AFICHER LA PLAYLIST COMPLÈTE (36 VIDÉOS)

AFICHER LA PLAYLIST COMPLÈTE (6 VIDÉOS)

AFICHER LA PLAYLIST COMPLÈTE (5 VIDÉOS)

AFICHER LA PLAYLIST COMPLÈTE (8 VIDÉOS)

COURS DE STAT À AGROCAMBUS

mathAgrocampus

S'ABONNER

CHAÎNES SIMILAIRES

se préparer pour une ...

S'ABONNER

PatR67

S'ABONNER

XLSTAT

S'ABONNER

Analyse de données avec FactoMineR

TOUT REGARDER

L'analyse de données avec R et FactoMineR. Comment faire une ACP, une ACM, une AFM ou encore une classification ? Quelles aides à l'interprétation ? Comment construire des graphiques lisibles ?

Analyse en composantes principales (ACP) Théorie et pratique

F. Husson

49:30

ACP avec le menu déroulant de FACTOMINER

F. Husson

18:13

ACP avec FACTOMINER


F. Husson


14:49





FACTOMINER^R

 **Factoshiny** : graphes interactifs
FactoInvestigate : rapports automatiques

 **missMDA** : données manquantes



FACTOMINER^R

R *Factoshiny* : graphes interactifs
FactoInvestigate : rapports automatiques

R *missMDA* : données manquantes



The **R** Journal



Journ
 April 2016, 11

FACTOMINER^R

Factoshiny : graphes interactifs
FactoInvestigate : rapports automatiques

missMDA : données manquantes



The **Journal**



Journal
 April 2016, 11

FACTOMINER^R



MOOC analyse de données multidimensionnelles

R *Factoshiny* : graphes interactifs
FactInvestigate : rapports automatiques

R *missMDA* : données manquantes



AgroCampus Ouest Analyse des données

Rechercher un cours François Husson

Accueil Tous les cours Analyse des données multidimensionnelles

Analyse des données multidimensionnelles

Voir le page de présentation du cours dans Studio

Les thématiques : **Mathématiques et statistiques** Sciences pour l'ingénieur

A propos du cours
 Cette 2ème édition augmentée du cours d'analyse de données multidimensionnelles débutera le 1er mars 2016.

Vous êtes inscrit à ce cours
 Voir le contenu du Cours

- Fin d'inscription
- Début du Cours
- Fin du cours
- Effort estimé



The **R** Journal



Journal of Statistical Software
 April 2016, Volume 70, Issue 1
 doi:10.18637/jss.v070.i01

FACTOMINER

MOOC analyse de données multidimensionnelles

Objectif : comprendre et savoir mettre en œuvre 5 méthodes d'analyse des données (ACP, AFC, ACM, classification et AFM)

Chaque semaine on trouve :

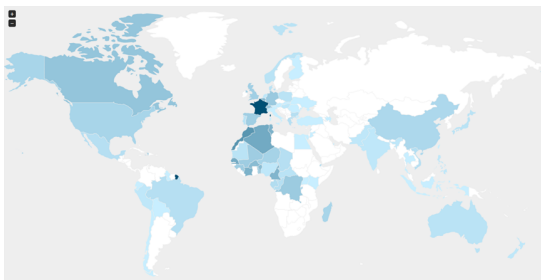
- des vidéos de cours : description de la méthode
- des quiz et exercices sur table
- une vidéo sur la mise en œuvre logicielle (FactoMineR)
- un exercice sur ordinateur
- (facultatif) des vidéos "Pour aller plus loin"

Egalement un forum et un Wiki

6 sessions en français, 4 en anglais :

- Depuis 2015 en FR et 2017 en GB sur la plateforme **FUN**
NOUVELLE SESSION en mars 2020 en français et en anglais

Quelle audience pour ce MOOC ?



Tous les ans :
 $\approx 5\,200$ FR
 $\approx 1\,800$ GB
112 pays
18 à 78 ans ($\bar{x} = 35$)
60% Master, 18% pHD

- Monde industriel et académique : Entreprises, Instituts de recherche (inra, inserm, ird...), Universités (France, Turquie, Tunisie, Zaïre...)
- Différents champs disciplinaires : biologie, économie, géographie, linguistique, etc.

⇒ Audience très variée avec des buts différents et des niveaux d'investissement différents

MOOC utilisé en pédagogie inversée

Organisation :

- module de 25h d'analyse de données en pédagogie inversée
- 192 étudiants de M1 d'Agrocampus – 3 enseignants (pour 10 heures)

MOOC utilisé en pédagogie inversée

Organisation :

- module de 25h d'analyse de données en pédagogie inversée
- 192 étudiants de M1 d'Agrocampus – 3 enseignants (pour 10 heures)

Contenu :

- mêmes cours, quiz, exercices, forum, wiki
- permanences pour discuter du cours et de leur projet
- évaluation par un quiz, un projet et l'évaluation de projet

MOOC utilisé en pédagogie inversée

Organisation :

- module de 25h d'analyse de données en pédagogie inversée
- 192 étudiants de M1 d'Agrocampus – 3 enseignants (pour 10 heures)

Contenu :

- mêmes cours, quiz, exercices, forum, wiki
- permanences pour discuter du cours et de leur projet
- évaluation par un quiz, un projet et l'évaluation de projet

Points positifs :

- étudiants plus autonomes
- questions-discussions avec les étudiants plus intéressantes
- effet positif des discussions et commentaires des apprenants (cours utile)
- plus de temps pour discuter et échanger sur leur projet
- correction par les pairs

MOOC utilisé en pédagogie inversée

Organisation :

- module de 25h d'analyse de données en pédagogie inversée
- 192 étudiants de M1 d'Agrocampus – 3 enseignants (pour 10 heures)

Contenu :

- mêmes cours, quiz, exercices, forum, wiki
- permanences pour discuter du cours et de leur projet
- évaluation par un quiz, un projet et l'évaluation de projet

Points positifs :

- étudiants plus autonomes
- questions-discussions avec les étudiants plus intéressantes
- effet positif des discussions et commentaires des apprenants (cours utile)
- plus de temps pour discuter et échanger sur leur projet
- correction par les pairs

Limites :

- certains étudiants préfèrent la douce tranquillité du présentiel
- ceux qui ne travaillent pas sont difficiles à détecter

Le MOOC selon moi

Ce que ça apporte

- Tout ce dont j'ai déjà parlé mais aussi ...
- une visibilité accrue (chaîne Youtube 1000 vues/j, MOOC \approx 130 000 vues)
- un tutoriel de référence très complet pour un package
- un guide complet pour un ensemble de méthodes
- une ouverture vers de nouveaux publics
- une incitation à repenser l'enseignement et la pédagogie

Le MOOC selon moi

Ce que ça apporte

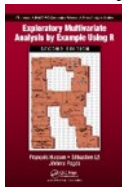
- Tout ce dont j'ai déjà parlé mais aussi ...
- une visibilité accrue (chaîne Youtube 1000 vues/j, MOOC \approx 130 000 vues)
- un tutoriel de référence très complet pour un package
- un guide complet pour un ensemble de méthodes
- une ouverture vers de nouveaux publics
- une incitation à repenser l'enseignement et la pédagogie

MAIS

- demande un fort investissement : 1h de cours \approx 20-30h !!!
Séquençage, quiz et exercices sont aussi chronophages
Les MOOC vieilliront-ils bien dans le temps ?

Livres

Analyse de données avec R (2^e ed)



- *F. Husson, S. Lê & J. Pagès (2016). 2^e édition, 240p.*
- ACP, AFC, ACM et classification
- Cours et nombreux exemples avec interprétation détaillées et lignes de code R

R pour la stat. et sc. des données



- *F. Husson, et al. (2018). 415p.*
- 1^{ère} partie : Concept, manipuler les données, visualiser les données, programmer
- 2^e partie : 29 méthodes illustrées par un exemple commenté

D'un package à un projet sur l'analyse de données

- 2003 : quelques programmes en S (Splus) : ACP, AFM
- avr 2006 : package FactoMineR (PCA, CA, MCA, MFA, etc.)
- avr 2007 : article sur FactoMineR (Journal of Statistical Software)
- avr 2007 : site internet
- nov 2007 : package RcmdrPlugin.FactoMineR (menu déroulant)
- juin 2009 & fév 2016 : Analyse de données avec R
- mai 2010 : package missMDA (gestion des données manquantes)
- jan 2011 & fév 2017 : Exploratory Multivariate Analysis by Example Using R
- avr 2011 : tutoriels vidéos
- sep 2011 : Google group
- nov 2012 : Análisis de datos con R
- mar 2013 : chaîne Youtube (vidéos de cours et didacticiel)
- juin 2013 : article sur MFACT (R Journal)
- fév 2014 : cours en ligne d'analyse de données (en local)
- fév 2015 : package Factoshiny (graphes interactifs)
- mar 2015-19 : MOOC d'analyse de données (FUN)
- juin 2015 : article sur CA-Galt (R Journal)
- avr 2016 : article sur missMDA (Journal of Statistical Software)
- Mar 2017-19 : MOOC Exploratory Multivariate Data Analysis (English)
- nov 2019 : package FactoMineR 2.0 (graphes ggplot)
- dec 2019 : package Factoshiny 2.0
- Mar 2020 : New session MOOC (FR or EN)

Un projet en analyse de données

Factoshiny : graphes interactifs
FactoInvestigate : rapports automatiques

missMDA : données manquantes



FACTOMINER



Retrouver le diaporama ici